Findings MLVR/DNN/RNN

Different models have been created and used ,whereas there were different findings per model in different set conditions. The first model that has been applied is the multivariate regression model. Essentially, multivariate analysis is a tool to find patterns and relationships between several variables simultaneously. It lets us predict the effect a change in one variable will have on other variables. At first glance the multivariate regression model is the simplest model to compute. To gain more insight in this model let's discuss the advantages and disadvantages of each model. To ensure that the results of the models were not influenced, the parameters that had the least correlation with each other.

Advantage

- It works on any dataset , gives information about the relevants of features
- It is a fast calculating model
- The MVLR is not a complicated model and works great do to its structural simplification
- Despite an apparent simplicity, they are very useful on a huge amount of features where better algorithms suffer from overfitting
- Great ability to identify outliers or anomalies
- Forecasting: Ability to determine the relative influence of predicter values to criteria values.

Disadvantages

- MVLR can suffer from multicollinearity, autocorrelation, heteroskedasticity
- Multicollinearity can increase the variance of the coefficient estimates and make the estimates very sensitive to minor changes in the model. The result is that the coefficient estimates are unstable.
- Can falsely conclude that a correlation is a causation.

DNN

The Deep Neural Network uses a variety of multiple layers of processing units and feature extraction and transportation. Each successive layer uses the output from the previous layer as input. This was the first simplified use of a NN. In the DNN there has been use of two hidden layers.

But as been stated before models have their advantages and disadvantages:

Advantages

- Reduces the need for feature engineering, one of the most time-consuming parts of machine learning practice.
- Is an architecture that can be adapted to new problems relatively are using techniques like convolutional neural networks, recurrent neural networks(see ref)

Disadvantages

- Requires a large amount of data ,deep learning is unlikely to outperform other approaches.
- Is extremely computationally expensive to train. The most complex models take weeks to train using hundreds of machines equipped with expensive GPUs.
- Do not have much in the way of strong theoretical foundation. This leads to the next disadvantage.

RNN

With RNN we've looked back three timesteps to predict one timestep ahead which in theory can be used to predict more timesteps ahead. The most probable this advantage from the RNN model that has been computed is that the results are hard to reproduce, it changes everytime the notebooks get run. At the

Advantage

- The main advantage of RNN over DNN is that RNN can model sequence of data (i.e. time series) so that each sample can be assumed to be dependent on previous ones. On the contrary, ANN can not model sequence of data. So, DNN is useful if only each sample is assumed to be independent of previous.
- It is possible to use same transition function f with same parameters at every time step.

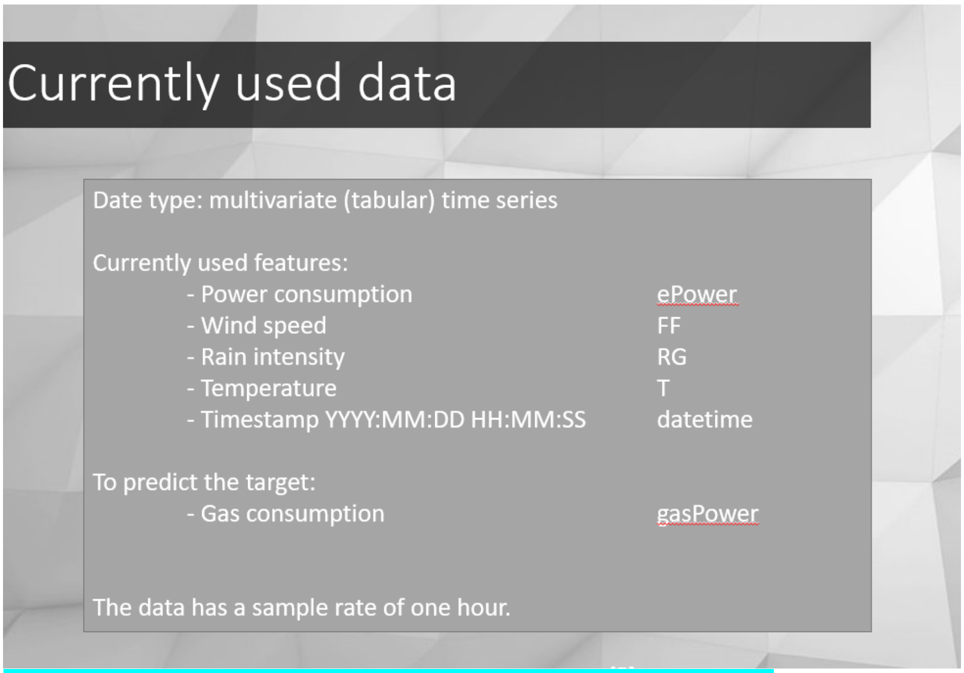
All RNNs have feedback loops in the recurrent layer. This lets them maintain information in 'memory' over time. But, it can be difficult to train standard RNNs to solve problems that require learning long-term temporal dependencies. This is because the gradient of the loss function decays exponentially with time (called the vanishing gradient problem). LSTM networks are a type of RNN that uses special units in addition to standard units. LSTM units include a 'memory cell' that can maintain information in memory for long periods of time. A set of gates is used to control when information enters the memory, when it's output, and when it's forgotten. This architecture lets them learn longer-term dependencies. GRUs are similar to LSTMs, but use a simplified structure. They also use a set of gates to control the flow of information, but they don't use separate memory cells, and they use fewer gates.

Disadvantages

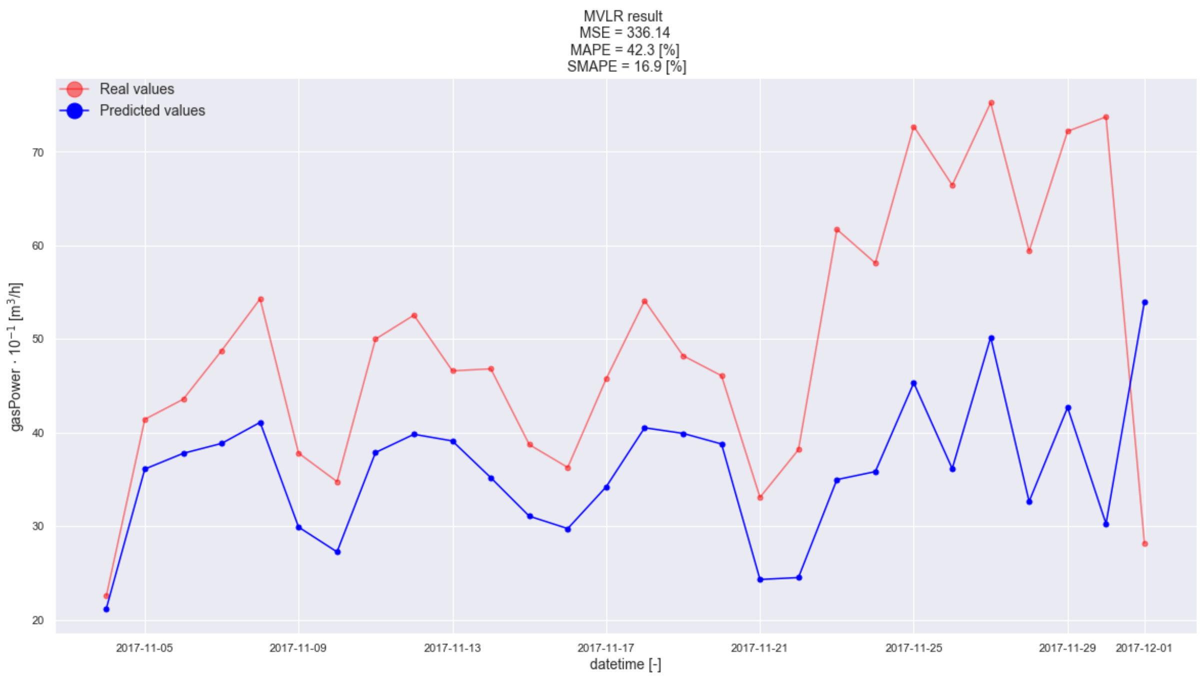
There are 2 main issues in Recurrent neural networks (RNNs) and they are related to difficulty in their training :

- Vanishing Gradient Problem: it is like the activation of one ANN is input to the next ANN (in time) where there are many such linked ANNs (in time). Actually the output activation is again fed as input to the same ANN. So it is like a replica of the same ANN in time steps. Thus there are N ANNs with same weights, but different inputs and they are linked via activations. Now when the gradients of the activation propagate from next ANN to previous one and so on, the gradients multiply due to chain rule of differentiation. So if the gradient is very less near to zero in one ANN, it is likely that the gradient at the previous Ann is also very small as they have the same weights. Also the gradients multiply and further reduce. This causes a chain of smaller gradients at each time step ANN.
- Similarly there is issue of increasing gradients at each step called as exploding gradients

It also has to be noted that an LSTM and GRU's are a type of RNN. LSTM networks are a type of RNN that uses special units in addition to standard units. LSTM units include a 'memory cell' that can maintain information in memory for long periods of time. GRUs are similar to LSTMs, but use a simplified structure.



Mention this in methodology on how we selected these features



If we compare the models to each other we notice that the MSE, MAPE and SMAPE are higher with the DNN model in comparison to the MVLR. With this being said the SMAPE limits outlier effects in both models. In a symmetric fashion, it also hides gross prediction errors as well. As seen in the graphs the MAPE scale in sensitive sensitive and should not be used when working with low-volume data.

If we take in to account that CNN and LSTM have a lower performance than RNN. Standard RNNs (Recurrent Neural Networks) suffer from vanishing and exploding gradient problems. LSTMs (Long Short Term Memory) deal with these problems by introducing new gates, such as input and forget gates, which allow for a better control over the gradient flow and enable better preservation of "long-range dependencies".

Meanwhile, LSTM has both cell states and a hidden states. The cell state has the ability to remove or add information to the cell, regulated by "gates". And because of this "cell", in theory, LSTM should be able to handle the long-term dependency (in practice, it's difficult to do so.)

https://www.researchgate.net/post/What_is_the_advantage_of_using_RNN_and_LSTM_over_traditional_methods_for_time_series_of_streaming_data

https://medium.com/@Anvesh525/pros-and-con-s-of-rnn-lstm-gru-cnn-recursive-nn-in-nlp-9771637d8e4a

https://www.altoros.com/blog/recurrent-neural-networks-classifying-diagnoses-with-long-short-term-memory/

conclusion: Have not written about CNN, Drunkschaler model is the best(see baldiri mail), RNN has higher performance than CNN(51%) MAPE and LSTM(Fancy RNN).

Not sure about values, values are contradictive to notebooks, so left spaces open

Models	**MSE**	**MAPE**	**SMAPE**	**Epochs**	**Seconds/epoch**
MVLR		42.3	16.9Hour		
DNN	0.65	57.1	19.2hour	30	4 micro
RNN					
CNN					
“Opschaler”					

Intro

Smartmeter systems in residential housing are becoming increasingly common due to trying to lower energy consumption. However energy-saving is still an important topic and many countries are still worried about energy related issues. The first and foremost source of energy is natural gas. Most residential houses use natural gas for heating, cooking and electricity generation. Researchers of the TU Delft have already made steps to eventually be able to predict energy consumption with the use of smart meter data. In this regard, to be able to accurately predict gas consumption with least parameters as possible can change the way people consume and help the world build a better and cleaner future.

Ultimately, accurate and cost effective wide-scale energy prediction is a vital step towards next-generation energy efficiency initiatives, which will require not only consideration of the methods, but the scales for which data can be distilled into meaningful information. In this regard, researchers have made great efforts to balance indoor thermal comfort and energy consumption. Among various research methods, data mining approaches have been extensively studied. These approaches can be further classified as energy consumption prediction, indoor air temperature prediction, and predictive control strategies.

Therefore, the ability to forecast and characterize building energy consumption is vital to implementing urban energy management and efficiency initiatives required to curb emissions. Advances in smart metering technology have enabled researchers to develop “sensor based” approaches to forecast building energy consumption that necessitate less input data than traditional methods.